

ニュースを選別する手間を軽減させる RSS リーダの提案

035164 高林 光

(指導教員 速水 治夫教授)

1 はじめに

Web サイトを要約するために作られた RSS と呼ばれるメタデータがある。RSS はサイトの更新情報を伝える手段であり、RSS リーダと呼ばれるアプリケーションを用いて、直接 Web サイトを閲覧することなく、サイトの更新状況を確認できる。

しかし、昨今の情報爆発により、更新情報だけでもその量は膨大になり、そこから必要な情報のみをえり抜く労力は、全て RSS リーダ利用者の負担となり、受信する RSS の量が増えるほど、こうした負担は大きな問題へと繋がる。

そこで本研究では、利用者のコンテンツ閲覧履歴から導き出した利用者自身の嗜好を用いて、収集した膨大な量の情報を取捨選択するアルゴリズムを RSS リーダに組み込む方法を検討した。

2 システム概要

本研究で試作した RSS リーダの概要を図 2.1 に示す。

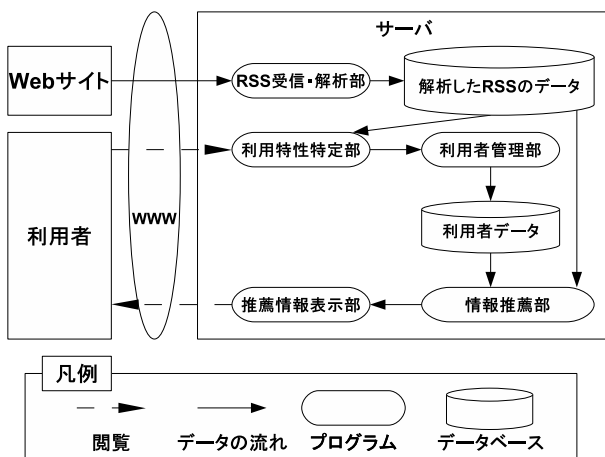


図 2.1 試作した RSS リーダの概要

本研究の RSS リーダでは、利用者の利用特性を分析し、利用者の嗜好に沿った情報を推薦表示する。そのため利用者は、Web ブラウザを用いて試作 RSS リーダが存在するサーバにアクセスし、利用者管理部で ID とパスワードを登録し、ログインする。

情報推薦部が、RSS 受信・解析部で格納された全ての RSS のコンテンツのタイトルと説明を利用者の利用特性と照らし合わせ、推薦情報表示部に結果を送り、利用者毎に異なったコンテンツを表示することで、利用者の嗜好に沿った情報の推薦表示を実現する。

3 評価

試作した RSS リーダを評価するため、適合率と再現率を算出し、評価値とする。再現率は、利用者の興味があるコンテンツをシステムがどれだけ網羅しているか、適合率は、利用者の興味があるコンテンツをシステムがどれだけ正確に抽出しているか、を表している。

システムの特性上、利用者がコンテンツを閲覧すればするほど、利用者の嗜好が顕著に表れてくるため、1つのコンテンツを閲覧する度に再現率と適合率を逐一算出し、それらの平均値を評価値とした。1名の利用者による、評価結果を以下の表 3.1 に示す。

表 3.1 評価結果

閲覧数	再現率	適合率
1	0.03	0.09
2	0.10	0.33
3	0.16	0.20
4	0.27	0.31
5	0.18	0.29
6	0.21	0.32
7	0.21	0.32
8	0.18	0.18
9	0.18	0.18
10	0.27	0.25
平均値	0.247	0.179

平均値は極めて低くなった。これは、利用者の利用特性を単語の出現頻度だけで決めていることに問題がある。

4 まとめ

評価結果より、問題は解決したとはいえないが、糸口を掴むことはできた。

今後、さらなるアルゴリズムの改良を行い評価基準の向上を図る。